Improving reference equations for cardiorespiratory fitness using multiplicative allometric rather than additive linear models: Data from the Fitness Registry and the Importance of Exercise National Database Registry

Running Title: An allometric reference equation for VO_{2max}

Authors: Alan M. Nevill[1], Jonathan Myers[2,3], Leonard A. Kaminsky[4], Ross Arena[5]

[1]Faculty of Education, Health and Wellbeing, University of Wolverhampton, Walsall Campus, Walsall, WS1 3BD, UK
[2]Division of Cardiology, VA Palo Alto Healthcare System, Palo Alto, CA
[3]Stanford University, Stanford, CA
[4]Fisher Institute of Health and Well-Being and Clinical Exercise Physiology Laboratory, Ball State University, Muncie, IN
[5]Department of Physical Therapy, College of Applied Sciences, University of Illinois at Chicago, Chicago, IL

Address for correspondence:

Professor Alan M. Nevill, Ph.D.
University of Wolverhampton,
Faculty of Education, Health and Wellbeing,
Walsall Campus
Gorway Road
Walsall, WS1 3BD
Tel: +44 (0)1902 322838
Fax: +44 (0)1902 322894
Email: a.m.nevill@wlv.ac.uk

Abstract

New improved reference equations for cardiorespiratory fitness have recently been published, using Data from the Fitness Registry and the Importance of Exercise National Database (FRIEND Registry). The new linear equation for VO_{2max} ($ml.kg^{-1}.min^{-1}$) was additive, derived using multiple-linear regression. An alternative multiplicative allometric model has also been published recently, thought to improve further the quality of fit. The purpose of the current study was to compare the accuracy and quality/goodness-of-fit of the linear, additive model with the multiplicative allometric model using the FRIEND database. The results identified that the allometric model out performs the linear model based on all model-comparison criteria. The allometric model demonstrates; 1) greater explained variance (R^2=0.645; R=0.803) vs. (R^2=0.62; R=0.79), 2) residuals that were more normally distributed, 3) residuals that yielded less evidence of curvature, 4) superior goodness-of-fit statistics i.e., greater maximum log-likelihood (MLL) and smaller Akaike Information Criterion (AIC) statistics, 5) less systematic bias together with smaller unexplained standard error of estimates. The Bland and Altman plots also confirmed little or no evidence of curvature with the allometric model, but systematic curvature (lack-of-fit) in the linear model. The multiplicative allometric model to predict VO_{2max} was;

$$VO_{2max}\ (ml.kg^{-1}.min^{-1}) = M^{-0.854} \cdot H^{1.44} \cdot \exp\ (0.424 - .346 \cdot (sex) - 0.011.age),$$

where M=body mass and H=height (R^2=0.645; R=0.803) and sex is entered as a [0,1] indicator variable (male=0 and female=1). Another new insight obtained from the allometric model (providing construct validity) is that the height-to-body-mass ratio is similar to inverse body mass index or the lean body mass index, both associated with leanness when predicting VO_{2max}. In conclusion adopting allometric

models will provide more accurate predictions of VO_{2max} ($ml.kg^{-1}.min^{-1}$) using more plausible, biologically sound and interpretable models.

Keywords: Curvature, power-function models, quality of fit, residuals.

Abbreviations and acronyms

BMI = body mass index
CPX = cardiopulmonary exercise test
CRF = cardiorespiratory fitness
CV = cardiovascular
CVD = cardiovascular disease
H = height
iBMI = inverse body mass index
KS = Kolmogorov-Smirnov
LBMI = lean body mass index
Ln = $\log_e$
M = body mass
RER = respiratory exchange ratio
VO_{2max} = maximal oxygen uptake

Conflicts of interest statement

The authors have no conflicts of interest

Introduction

Cardiorespiratory fitness (CRF), which can be accurately assessed by direct measurement of maximal oxygen consumption (VO_{2max}), is a well-established and robust indicator of cardiovascular (CV) health, as well as a valuable predictor of all-cause mortality.[1,2,3,4] However, despite its importance, direct measurements of VO_{2max} in epidemiological or population studies are rare primarily due to feasibility issues related to the time requirement to administer the test and the lack of having this measure routinely measured in clinical practice. As such, estimating or predicting CRF (i.e., VO_{2max}) has emerged as an attractive alternative; research to refine/improve VO_{2max} prediction models is ongoing.

Various prediction models using additive linear equations have been published recently, some include estimates of physical activity (PA)[5] while others do not. [6] However, Nevill and Cook[7] highlighted a number of concerns with these linear, additive models. Firstly, the models suggest linear associations with all key predictors such as age and body mass (M). However, there is strong evidence, certainly from the findings reported by Myers et al.[6] (see Figures 1 and 2), that curvature exists suggesting that at least one of these associations is likely to be non-linear.

The second concern is the absence of a body weight/mass term in the linear model proposed by Nes et al.,[5] and the absence of a height term in the model reported by Myers et al.[6] Astrand and Rodahl[8] in their Figure 9-4 on page 400 and Nevill et al., [9] reported a strong negative association between VO_{2max} ($ml.kg^{-1}.min^{-1}$) and body mass. This is because absolute VO_{2max} ($l.min^{-1}$) scales to, or is associated with, body mass ($M^{0.67}$). Thus, when researchers calculate VO_{2max} ($ml.kg^{-1}.min^{-1}$)

relative to body mass by dividing VO_{2max} ($l.min^{-1}$) by body mass (M), the resulting ratio over-scales, leaving VO_{2max} ($ml.kg^{-1}.min^{-1}$) proportional to $M^{-0.33}$. This **non-linear** association with body mass was not considered by Nes et al.[5] or Myers et al.[6] By incorporating a power-function body-mass term as a predictor variable in their models, the result is likely to explain some of the curvature described above and, at the same time, improve the accuracy of the model. Furthermore, neither of the linear models proposed by Nes et al.[5] nor Myers et al.[6] incorporated height as a predictor variable. When height was included as a predictor variable in the allometric model to predict VO_{2max} ($ml.kg^{-1}.min^{-1}$) proposed by Nevill and Cook,[7] a strong association was detected ($P<0.001$).

Another concern with these fitted linear additive models is the fact that the residuals from both models are unlikely to be: a) normally distributed; and b) homoscedastic (errors remain constant throughout the range of measurements; see **Figure 2** in Myers et al.[6]). If the residuals demonstrate a lack of normality and heteroscedasticity, then the validity of the models (i.e., the statistical significance of the estimated parameters) will be questionable. The alternative approach proposed by Nevill and Cook,[7] incorporating proportional, multiplicative allometric models, was found to overcome or at least reduce many of these problems; in particular, improving the normality and heteroscedasticity observed in the residuals. For a brief and concise history of allometric modeling, see Winter and Nevill.[10]

The purpose of the current study was to fit the same linear, additive model adopted by Myers et al.[6] to directly measure VO_{2max} ($ml.kg^{-1}.min^{-1}$) data from the "Fitness Registry and the Importance of Exercise: A National Data Base" (FRIEND) Registry, to compare the original linear additive model with an alternative, proportional allometric model.[7] The comparison assessed whether the latter

provides: 1) a superior quality of fit (using R^2, maximum log-likelihood and AIC criterion); 2) more normally distributed residuals; 3) less explainable bias and smaller unexplained standard deviation of differences (standard error of the estimate); and 4) a more plausible, biologically sound and interpretable model.

Methods

The procedures used for acquiring and managing the data for the FRIEND registry have been previously reported.[11] In brief, laboratories determined by the advisory board to use valid and reliable calibration and cardiopulmonary exercise test (CPX) procedures administered by experienced personnel were invited to be considered for inclusion in the FRIEND Registry. Although there were some variations in laboratory equipment, protocols, and procedures defining VO_{2max}, the characteristics of all participating CPX laboratories are consistent with recommendations outlined in recently published guidelines.[12,13] Local institutional review board approval for participation in the FRIEND Registry was obtained by each participating CPX laboratory to submit de-identified, coded data to the data coordinating center at Ball State University, which then forwarded these data to the core CPX laboratory housed at the University of Illinois at Chicago. Institutional review board approval for the core CPX laboratory was also obtained at the University of Illinois at Chicago. Data from each CPX laboratory were reviewed for uniformity and to ensure data were within expected normal ranges by both the coordinating center and the core laboratory prior to merging into the FRIEND database.

Study sample

The study sample included 7759 subjects (4601 men and 3158 women, mean 46 ± 13 years) from ten participating CPX laboratories with geographical representation from Indiana, Louisiana, North Carolina, Oregon, Pennsylvania, Tennessee, and Texas. For inclusion, subjects were required to meet the following criteria: 1) age $\geq$20 years; 2) a maximal exercise test performed on a treadmill; and 3) a peak respiratory exchange ratio (RER) ≥1.00. The indications for the exercise tests were determination of fitness level before entry into an exercise program or for a research study. Laboratories provided data on individuals who at the time of the test were without known CV disease (CVD;i.e., coronary artery disease, peripheral vascular disease, or heart failure) or chronic obstructive pulmonary disease. Any subject identified as having a pre-existing diagnosis of CVD or pulmonary disease was excluded.

Statistical methods

The additive, linear model proposed by Myers et al.[6] is given by

$$VO_{2max}\ (ml.kg^{-1}.min^{-1}) = a + b_1 \cdot age + b_2 \cdot M + b_3 \cdot H + \varepsilon, \qquad \text{(Eq. 1)}$$

where M=mass, H=height (note that H=height has been added for comparative purposes) and ε is an additive error term that is assumed to be both normally distributed and homoscedastic (remains constant throughout the range of observations). The intercept “a” was allowed to vary with sex.

An alternative multiplicative model with allometric body size components originally proposed by Nevill and Holder[14] and subsequently reported by Nevill and Cooke[7] is given by

$$VO_{2max}\ (ml.kg^{-1}.min^{-1}) = M^{k_1} \cdot H^{k_2} \cdot \exp(a + b_1 \cdot age) \cdot \varepsilon, \qquad \text{(Eq. 2)}$$

where M=mass, H=height and 'ε' is a multiplicative, error ratio that assumes the error will be in proportion to VO_{2max} ($ml.kg^{-1}.min^{-1}$).

The model (Eq. 2) can be linearized with a log transformation (using $Ln=log_e$). A linear regression analysis on $Ln(VO_{2max})$ can then be used to estimate the unknown parameters in the log transformed model i.e., the transformed model (Eq3) is now additive that conforms with the assumptions associated with ordinary least squares and ANOVA:

$$Ln(VO_{2max}) = k_1 \cdot Ln(M) + k_2 \cdot Ln(H) + a + b_1.age + Ln(\varepsilon), \quad \text{(Eq. 3)}$$

where the residual errors $Ln(\varepsilon)$ are assumed to be normally distributed and homoscedastic and the intercept "a" is allowed to vary with sex.

Normality was assessed using the Ryan-Joiner and Kolmogorov-Smirnov tests. The Ryan-Joiner statistic measures how well the data follow a normal distribution by correlating the association between the measured data and the calculated normal scores. If the correlation coefficient is near 1, the population is likely to be normal. Larger values for the Kolmogorov-Smirnov statistic (KS) indicate that the data do not follow the normal distribution.

Cross-validation

To assess the validity or success of the two competing models, we adopted the same cross-validation proposed by Nes et al.,[5] by splitting the FRIEND data into two independent groups using a random split (80:20). We used the 80% sample to predict the competing two models and the 20% sample to test/validate the models. This was achieved by predicting the VO_{2max} of the 20% sample using the model derived from the 80% prediction models. The success of cross-validation process

was then assessed by comparing the measured VO_{2max} of 20% validation sample with the predicted VO_{2max} scores using correlations and Bland and Altman's limits of agreement[15] extended to assess systematic bias due to sex and age groups (using a two-way ANOVA), as well as the usual standard deviation of differences (The standard error of the estimate based on the ANOVA's residual mean-square errors). Bland and Altman plots are also reported and compared.

Results

Additive linear models

Fitting a similar linear model to that reported by Myers et al.,[6] plus an additional term for height (H), improves the fit as follows,

$$VO_{2max} = 41.38 - 10.88 \cdot (sex) - 0.378 \cdot age - 0.310 \cdot M + 0.227 \cdot H, \quad \text{(Eq. 4)}$$

where sex is entered as a [0,1] indicator variable (male=0 and female=1). The R^2 was 0.632 (R=0.795) with the standard error of estimate being 6.96 ($ml.kg^{-1}.min^{-1}$). Note the original fit reported by Myers et al.[6] (excluding height) was $R^2 = 0.62$ (R=0.79) and the standard error of estimate was 7.2 ($ml.kg^{-1}.min^{-1}$).

When the residuals were saved and plotted against the predicted values (fits) as part of the usual model assessment diagnostics, the additive linear model demonstrates a clear lack of fit (evidence of curvature and heteroscedasticity) as seen in **Figure 1a**. The Ryan-Joiner statistic was r=0.996 and the Kolmogorov-Smirnov statistic KS=0.032, indicating that the residuals were not normally distributed, P<0.01.

■ Figure 1a and 1b about here--

In an attempt to explain the curvature seen in **Figure 1a**, we examined the associations between VO_{2max} and the three predictor variables - body mass, height and age. Results suggest that the curvature was predominately due to the association between VO_{2max} and body mass (see **Figure 2**).

■ Figure 2 about here--

Multiplicative allometric models

Fitting the multiplicative allometric model to the FRIEND's data using Eq. 3, we obtained,

$$Ln(VO_{2max}) = -.854 \cdot Ln(M) + 1.44 \cdot Ln(H) + .424 - .346 \cdot (sex) - .011.age, \quad \text{(Eq. 5)}$$

where sex is entered as a [0,1] indicator variable (male=0 and female=1). The R^2 was 0.678 (R=0.824). (The R^2 for the linear additive model Eq. 4 was 0.632 (R=0.795)).

In contrast to the additive linear model, when the residuals were saved and plotted against the predicted values, the log-transformed multiplicative allometric model demonstrates an acceptable fit (little or no evidence of curvature nor heteroscedasticity), see **Figure 1b**. The Ryan-Joiner statistic was r=0.997 and the Kolmogorov-Smirnov statistic KS=0.028, P<0.01. Despite the fact that both statistics indicate that the residuals saved from the allometric model were not normally distribution they were closer to normality than the residuals saved using the linear additive model (Eq. 4).

Taking anti-logs of (Eq. 5), we obtain the multiplicative allometric model to predict VO_{2max} ($ml.kg^{-1}.min^{-1}$) as follows;

$$VO_{2max} = M^{-0.854} \cdot H^{1.44} \cdot \exp(0.424 - .346 \cdot (\text{sex}) - 0.011.\text{age}), \quad \text{(Eq. 6)}$$

where sex is entered as a [0,1] indicator variable (male=0 and female=1). The R^2 (measured vs. the predicted VO_{2max} using Eq. 6) was 0.645 (R=0.803).

Comparing the goodness-of-fit

Since the models are not nested or hierarchical, a direct comparison between two competing model forms (linear vs allometric) is not possible using traditional criteria such as the residual sum-of-squares, the standard error and the coefficient of determination (R^2). However, Nevill and Holder[14] and Nevill et al.[16] chose the maximum likelihood criterion and the Akaike Information Criteria (AIC) as their standard criterion of model assessment (quality of fit) that does not require the competing models to be either nested or hierarchical.

A simple modification of the maximum log likelihood criterion is able to produce the Akaike Information Criteria (AIC= -2×(maximum log-likelihood) + 2×(number of parameters fitted)), see goodness-of-fit data from both the linear and allometric models (**Table 1**).

■ Table 1 about here--

Cross-validation assessment

The FRIEND's dataset was randomly split into two independent samples: the prediction model sample =6214 (80.1%) and the validation/test sample = 1545 (19.9%).

The results from the cross-validation assessment (using the 20% validation sample) found the correlations between the measured VO_{2max} and the predicted VO_{2max} were r=0.796 (linear) and r=0.809 (allometric). The differences between the measured and

predicted VO_{2max} (bias) were assessed using Bland and Altman's Limits of Agreements, extended to incorporate "explainable" bias due to sex and age groups. The bias (measured – predicted VO_{2max}) from linear and allometric prediction models were assessed using a two-way ANOVA. The mean bias by age and sex can be seen in **Figures 3a and 3b**, respectively. Note that the "explainable" variance in bias from the linear additive model using the two-way ANOVA was R^2=1.9%. The equivalent "explainable" variance in bias from the allometric-model using the two-way ANOVA was R^2=1.2%.

- Figure 3a and 3b about here--

Having explained these "systematic" biases due to age and sex (using ANOVA), the remaining unexplained standard deviation of differences (standard error of estimate) were ±6.87 ($ml.kg^{-1}.min^{-1}$) (linear model) and ±6.71 ($ml.kg^{-1}.min^{-1}$) (allometric model) (based on the ANOVA's mean-square errors). The Bland and Altman plots (differences vs means) for the linear and allometric prediction models are given in **Figures 4a and 4b**, respectively.

- Figure 4a and 4b about here--

Discussion

The answer to the question "can we improve the reference equation for normal standards for VO_{2max}, using multiplicative allometric rather than additive linear models?" would appear to be categorically yes. The allometric model (Eq. 2) performs better than the linear model (Eq.1), originally proposed by Myers et al.[6] based on ALL model-comparison criteria: 1) The explained variance. The explained variance was greater using the allometric models, using either the log-transformed model (R^2=0.678; R=0.824) or the multiplicative allometric model (R^2=0.645;

R=0.803) compared with the linear models (excluding height) (R^2 = 0.62; R=0.79) or (including height) (R^2=0.632; R=0.795); 2) Tests of normality: Both the Ryan-Joiner and Kolmogorov-Smirnov tests indicate that the residuals saved from the linear and allometric models were not normally distributed. However, the residuals from the allometric model were closer to a normal distribution than the residuals saved from the linear model. However, the Ryan-Joiner statistics (Q-Q plot correlations) for the linear and allometric models were 0.996 and 997 respectively, resulting in arguably acceptable linearity; 3) The residuals vs the predicted values (fits) plots. Evidence of curvature (lack-of-fit) was observed with the linear model in **Figure 1a**. No such evidence was apparent with the allometric model in **Figure 1b**; 4) The goodness of fit assessed using the maximum log-likelihood (MLL) and the Akaike Information Criterion (AIC). The MLL was greater, and the AIC was smaller with the allometric model compared with the linear additive models (**see Table 1**).; 5) The cross-validation assessment. Based on the validation sample, the correlation between the measured VO_{2max} and the predicted VO_{2max} was r=0.796 (linear) and r=0.809 (allometric). The bias was less evident from the allometric model (R^2=1.2% compared with R^2=1.9% with the linear model) see **Figure 3b** compared with the linear model **Figure 3a**. The Bland and Altman plots confirmed a lack-of-fit (curvature) in the linear model but little evidence of curvature with the allometric model.

Perhaps the most persuasive argument for choosing the allometric model comes from the lack-of-fit/curvature observed in the three figures based on linear models: **Figure 1a, Figure 2** and **Figure 4a**. The same curvature was also observed in **Figures 1** and **2** reported by Myers et al.,[6] In effect, the linear model systematically underestimates VO_{2max} ($ml.kg^{-1}.min^{-1}$) among both lighter and heavier

participants but overestimates VO_{2max} ($ml.kg^{-1}.min^{-1}$) among the more "average" body mass participants. This might explain the negative bias of 70 year-old male participants (see **Figure 3a**) who are relatively light ("average" mass) having "survived" to the age of 70, and also the negative bias of a large, homogeneous group of 30 year-old female participants (see **Figure 3a**) who have maintained an "average" body mass especially compared to their male counterparts.

Two other helpful new insights were obtained from the fitted multiplicative allometric model (Eq. 6). The first comes from observing that the fitted mass (M) and height (H) exponents have opposite signs. The resulting product $H^{1.44} \cdot M^{-0.854}$ can be expressed as a ratio ($H^{1.44}/M^{0.854}$) or $(H^{1.69}/M)^{0.854}$ not dissimilar to the inverse BMI ($iBMI=H^2/M$). This index has a sound biological interpretation in terms of its association with cardiorespiratory fitness [i.e., VO_{2max} ($ml.kg^{-1}.min^{-1}$)]. The inverse BMI[17] also known as the Lean Body Mass Index (LBMI)[18], is a measure of leanness. Clearly having a greater lean body mass (LBMI) should be strongly and positively associated with VO_{2max} ($ml.kg^{-1}.min^{-1}$), thus providing further biological support (and construct validity) for the allometric model (Eq. 6).

The second helpful new insight obtained from the allometric model (Eq. 6) comes from the negative decline in VO_{2max} with age. The linear model incorporates the age decline as a negative linear term that would theoretically predict a negative VO_{2max} for very old participants. The allometric model incorporates the negative age decline within an exponential term, thus ensuring the prediction of VO_{2max} remains positive irrespective of the participants' age.

We acknowledge that the current study is not without limitations. Although we have been able to demonstrate the benefits of the allometric model for VO_{2max},

obtaining similar results to those reported by Nevill and Cooke,[7] further work is required. For example, future research should attempt to explain the systematic bias observed in **Figure 3b**, in particular the negative bias observed in the 70 year-old age groups.

In summary, the quality of fit associated with predicting VO_{2max} ($ml.kg^{-1}.min^{-1}$) using the allometric model was superior to the linear additive models, based on all model-comparison criteria. As seen in **Figure 2**, fitting a linear additive model will systematically over-estimate the VO_{2max} for participants who are either very light or very heavy, but under-estimate VO_{2max} for participants who have "average" body mass. The allometric model also identified a stature-to-body-mass ratio, very similar to LBMI or inverse BMI, known to be associated with leanness, a new insight that leads to a more plausible, biologically sound and interpretable model when predicting VO_{2max} ($ml.kg^{-1}.min^{-1}$).

Acknowledgments FRIEND Consortium Contributors: Ball State University (Leonard Kaminsky), Brooke Army Medical Center (Kenneth Leclerc), Cone Health (Paul Chase), Johns Hopkins University (Kerry Stewart), Pennington Biomedical Research Center (Timothy Church), Southern Connecticut State University (Robert Axtell), Taylor University (Erik Hayes), University of Illinois–Chicago (Jacob Haus), University of Tennessee, Knoxville (David Bassett).

References

1. Ross R, Blair SN, Arena R, et al. Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign: a scientific

statement from the American Heart Association. Circulation 2016;134(24):e653-e699.

2. Ozemek C, Laddu DR, Lavie CJ, et al. An update on the role of cardiorespiratory fitness, structured exercise and lifestyle physical activity in preventing cardiovascular disease and health risk. *Prog Cardiovasc Dis* 2018;61:484-490.
3. Kaminsky LA, Arena R, Ellingsen Ø, et al. Cardiorespiratory fitness and cardiovascular disease: the past, present, and future. *Prog Cardiovasc Dis* 2019;62:86-93.
4. Imboden MT, Harber MP, Whaley MH, et al. The association between the change in directly measured cardiorespiratory fitness across time and mortality risk. *Prog Cardiovasc Dis* 2019;62:157-162.
5. Nes BM, Janszky I, Vatten LJ, et al. Estimating VO2peak from a Non-exercise Prediction Model: The HUNT Study, Norway. *Med. Sci. Sports Exerc.* 2011;43:2024–30.
6. Myers, J., Kaminsky, L. A., Lima, R., et al. A reference equation for normal standards for VO2 max: analysis from the Fitness Registry and the Importance of Exercise National Database (FRIEND Registry). *Prog Cardiovasc Dis* 2017:*60*:21-29.
7. Nevill, A.M. and Cooke, C.B. The Dangers of Estimating VO2max Using Linear, Non-exercise Prediction Models. *Med. Sci. Sports Exerc.* 2017;**49**:1036–1042 <https://journals.lww.com/acsm-msse/Fulltext/2017/05000/The_Dangers_of_Estimating_V_O2max_Using_Linear,.22.aspx>.

8. Astrand P-O and Rodahl K: Textbook of work physiology, 3rd ed. New York: McGraw-Hill; 1986.
9. Nevill AM, Ramsbottom R, and Williams C. Scaling physiological measurements for individuals of different body size. *Eur J Appl Physiol.*1992;65: 110-7.
10. Winter E.M. and Nevill A.M. Scaling: Adjusting for differences in body size. In "Kinanthropometry and Exercise Physiology Laboratory Manual for Tests, Procedures and Data, 3rd ed." (eds) R. G. Eston and T. Reilly, Abingdon, Oxon, Routledge; 2009;300-314 p.
11. Kaminsky LA, Arena R, Beckie TM, et al. The importance of cardiorespiratory fitness in the United States: the need for a national registry. A policy statement from the American Heart Association. Circulation. 2013;127(5):652-662.
12. Balady GJ, Arena R, Seitsema K, et al. A clinician's guide to cardiopulmonary exercise testing. A scientific statement from the American Heart Association. Circulation. 2010;122: 191-225.
13. Myers J, Arena R, Franklin B, et al. Recommendations for clinical exercise laboratories: a scientific statement from the American Heart Association. Circulation. 2009;119(24):3144-3161.
14. Nevill AM and Holder RL. Modelling maximum oxygen uptake - A case study in non-linear regression model formulation and comparison. *Journal of the Royal Statistical Society, Series C.* 1994; 43: 653-66.
15. Bland, JM and Altman, D. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 1986;*327*(8476): 307-310.

16. Nevill AM, Holder RL, Baxter-Jones A, et al. Modeling developmental changes in strength and aerobic power in children. *J Appl Physiol.* 1998;84: 963-70.
17. Nevill AM, Stavropoulos-Kalinoglou A., Metsios G.S., et al. Inverted BMI rather than BMI is a better proxy for percentage of body fat. *Ann Hum Biol.* 2011; 38: 681-4.
18. Nevill AM and Holder RL. Body Mass Index; a measure of fatness or leanness? *The British Journal of Nutrition*, 1995;73: 507-516.

Legends to Tables

Table 1. The maximum log-likelihood (MLL) and Akaike Information Criterion (AIC) together with the number of fitted parameters for the competing models to predict VO_{2max}

Legends to figures

Figure 1. The association between the residuals and the predicted values (fits) saved from fitting 1a) the additive linear model (Eq. 1), and 1b) the log-transformed multiplicative allometric model (Eq. 5)

Figure 2. The association between VO_{2max} ($ml.kg^{-1}.min^{-1}$) and body mass (kg) for female participants (linear R^2=0.316; power function R^2=0.374)

Figure 3. The mean bias (measured – predicted VO_{2max}) from 3a) the linear prediction model, and 3b) the allometric prediction model, assessed using a two-way ANOVA (sex and age group)

Figure 4. The Bland and Altman plot (differences vs means) using the validation sample, data from 4a) the linear model and 4b) the allometric model to predict VO_{2max}

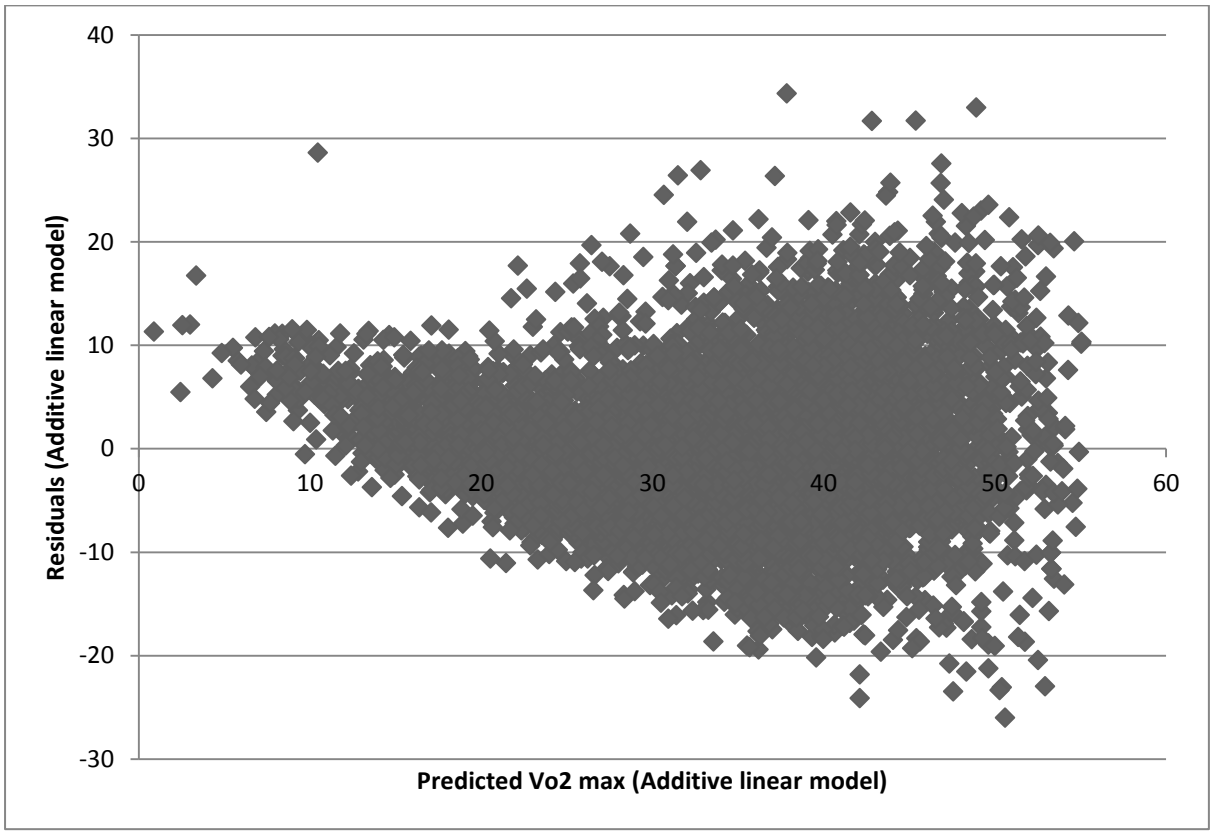
Table 1.

Competing models	MLL (N par)	AIC
Additive linear model	-25919.01 (4)	51846
Log-transformed allometric model (Eq. 3)	-25116.57 (4)	50241

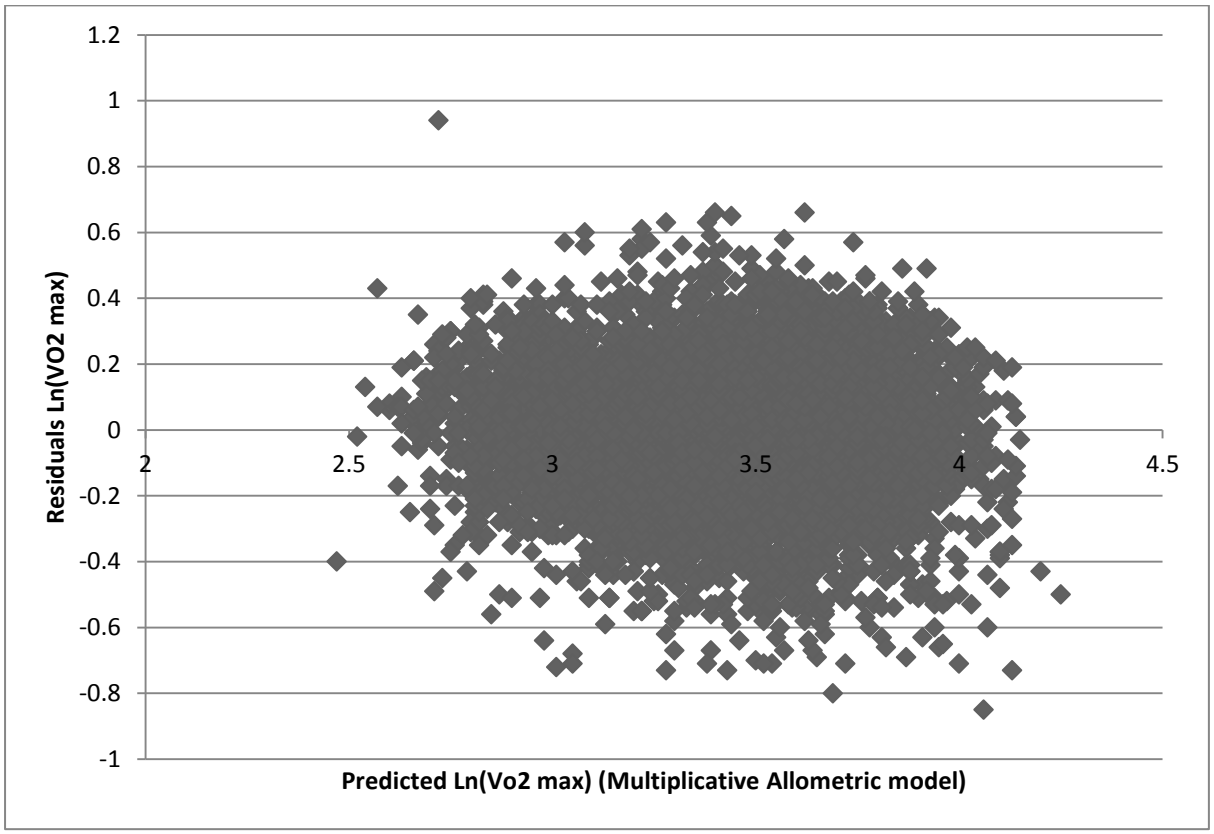
N par =number of fitted parameters

Note that the best model is the one with the greatest MLL (i.e., least negative) and/or the smallest AIC

Figure 1a.


40
30
20
10
0
-10
-20
-30
0
10
20
30
40
50
60
Residuals (Additive linear model)
Predicted Vo2 max (Additive linear model)

448 **Figure 1b**

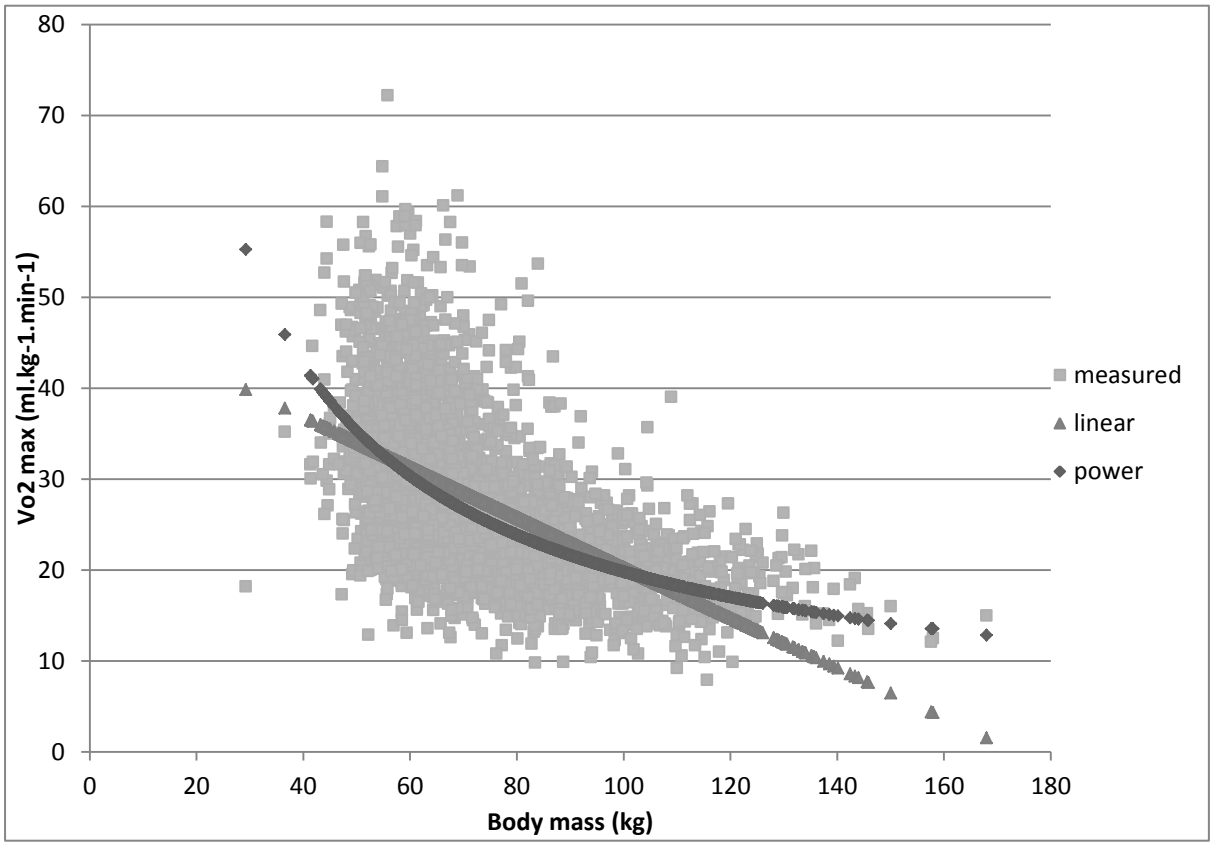


449

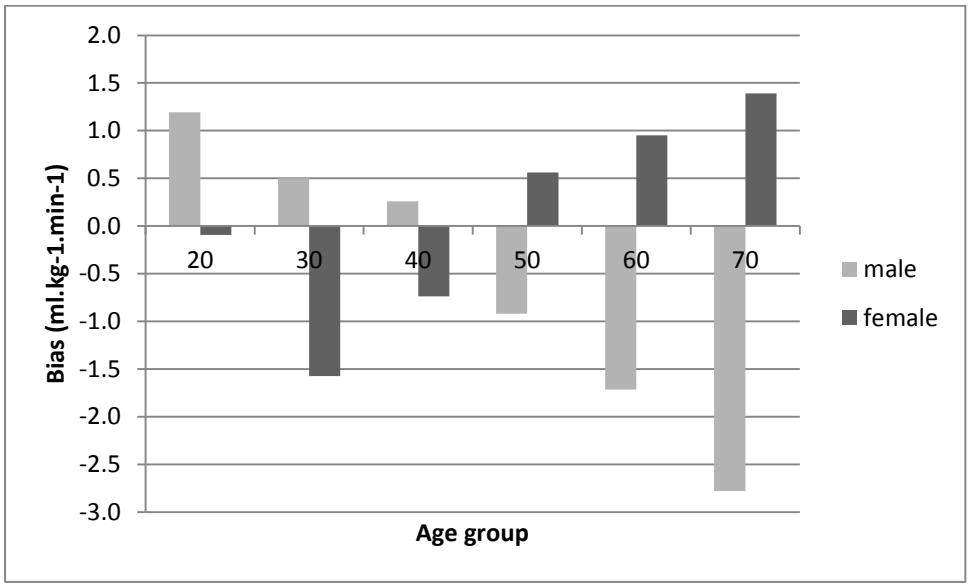
450

451

Figure 2.

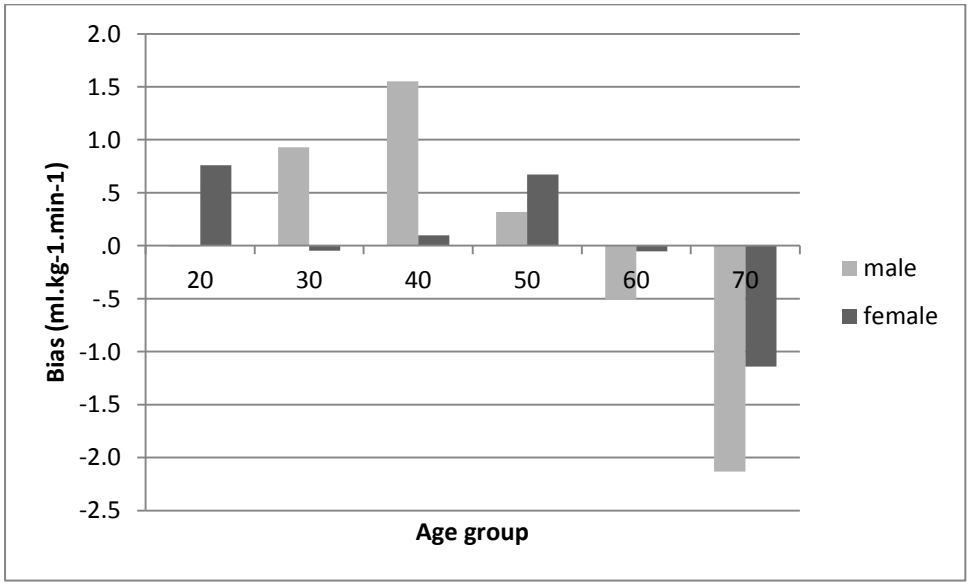

80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
120
140
160
180
Vo2 max (ml.kg-1.min-1)
Body mass (kg)
measured
linear
power

455 **Figure 3a**



456

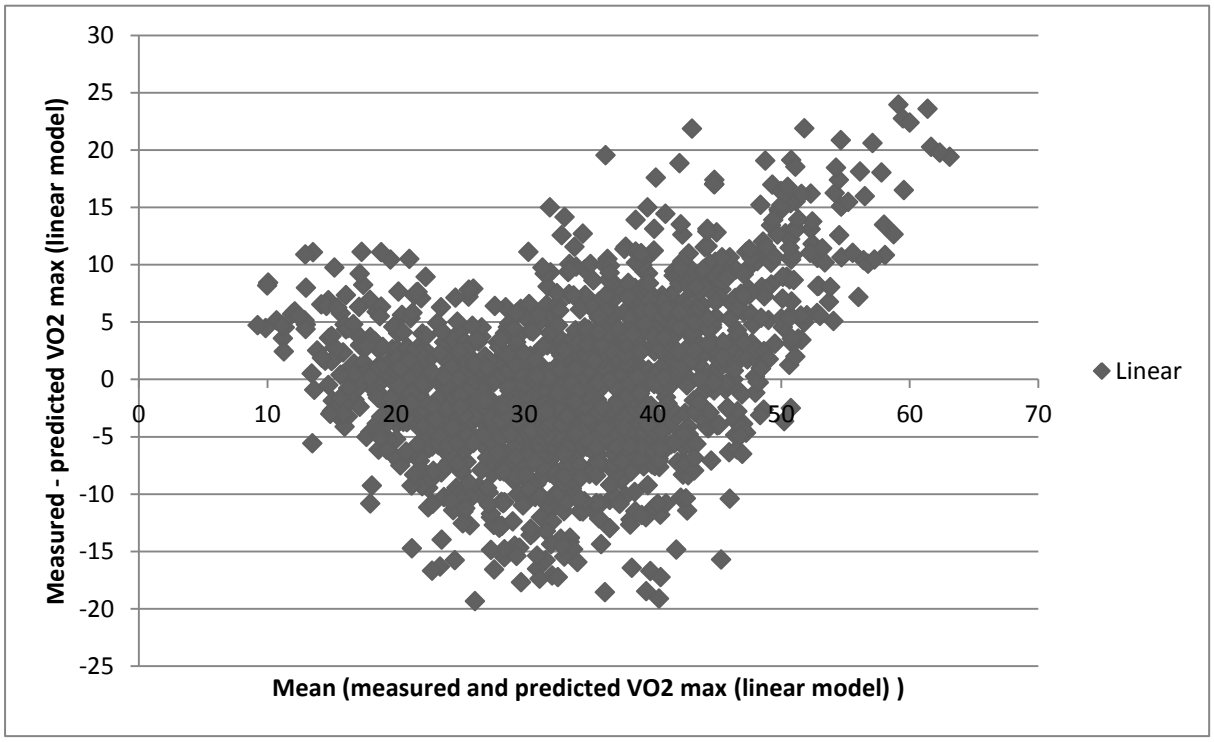
457 **Figure 3b.**



458

459

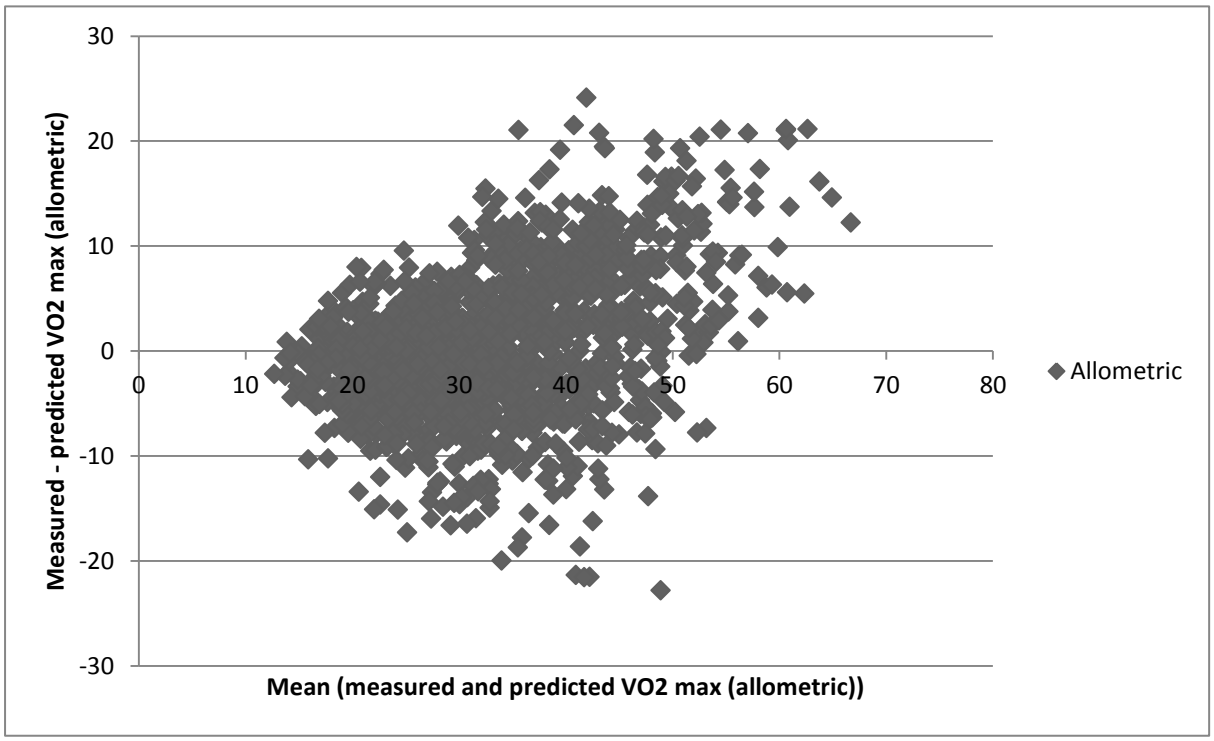
460 **Figure 4a.**



461

462

463 **Figure 4b.**



464